

Article

Modeling of Atmospheric Pollution in Urban and Rural Sites Using a Probabilistic and Objective Approach

Francisco J. Moral ^{1,*}, Francisco J. Rebollo ¹, Pablo Valiente ² and Fernando López ¹

¹ Departamento de Expresión Gráfica, Universidad de Extremadura, Avda. de Elvas, s/n, 06006 Badajoz, Spain; frebollo@unex.es (F.J.R.); ferlopez@unex.es (F.L.)

² Departamento de Química Analítica, Universidad de Extremadura, Avda. de Elvas, s/n, 06006 Badajoz, Spain; valiente@unex.es

* Correspondence: fjmoral@unex.es; Tel.: +34-924-289-600

Received: 26 June 2019; Accepted: 19 September 2019; Published: 25 September 2019



Abstract: Atmospheric pollution is affected by different individual pollutants (IP) and climatic factors (CF). In this work, the formulation of the Rasch model is proposed to get representative measures of atmospheric pollution in two urban locations, Badajoz and Cáceres, and one rural site, the Monfragüe Park (Southwest Spain). After applying the Rasch methodology, the ranking of all days was obtained according to their atmospheric pollution level and the influence on the environmental deterioration of each IP and CF (NO₂, NO, SO₂, O₃, CO, benzene, PM₁₀, precipitation, relative humidity, solar radiation, air temperature, and barometric pressure). The most influential items on atmospheric pollution are the O₃ and the CF, mainly the lack of precipitation and those related to ozone generation (air temperature and solar radiation). Other IP exert a lower influence at both urban locations, being irrelevant at the Monfragüe Park. Unexpected behaviors of the CF or IP can be also analyzed.

Keywords: atmospheric pollutant; climatic factors; rasch model; latent variable

1. Introduction

Monitoring studies of atmospheric pollution have become more important in the recent years, from local to planetary scale, because of the changes in human activities and their effects on climate and public health. The progressive degradation of nature, because of pollution, has as a consequence that people are demanding a less aggressive way of life for the environment, claiming clean industries, ecological produces, etc. Moreover, citizens demand to their governing class different measures to benefit the environment where they live, with the aim of favoring a better life quality.

The main air pollutants related to adverse effects are carbon monoxide (CO), nitrogen oxides (NO_x, consisting of NO and NO₂), ozone (O₃), sulfur dioxide (SO₂), volatile organic compounds (VOCs, including, for example, benzene and toluene), and particulate matter (usually considering PM₁₀, i.e., particulates with diameters lower than 10 μm). CO and NO are generated during the incomplete combustion of carbon-containing fuels, and NO₂ during more complete combustion conditions, mainly from traffic emissions, industrial activities and fuel-fed heating. Some VOCs are released as a result of human activity; for example, from evaporation of liquid fuels, vehicle exhaust, and from solvents in paints, which are known as anthropogenic sources. Moreover, significant amounts of VOCs are also emitted from vegetation, including urban forest and landscapes, agricultural crops, and natural communities. The VOCs emitted by plants are known as biogenic sources. Consequently, the total VOCs volume in the atmosphere is the sum of anthropogenic and biogenic VOCs. O₃ is a secondary pollutant produced by photochemical reactions of primary pollutants, NO_x and VOCs, collectively

called ozone precursors, enhanced by favorable meteorological conditions (high temperatures and strong solar radiation). The production and changes in the concentration of O_3 are critically dependent on the levels of precursors. The O_3 cycle starts with the emissions of NO_x into the atmosphere from a variety of biogenic and anthropogenic sources. The NO is rapidly converted in the atmosphere by the existing O_3 into NO_2 and O_2 . In the presence of sunlight, the NO_2 is photo-dissociated back into NO and atomic oxygen (O). Finally, the O recombines with O_2 to form O_3 . There are other species in the atmosphere that convert NO to NO_2 without destroying ozone. These are primarily VOCs, also called reactive hydrocarbons. The main atmospheric reactions of this class of compounds are complex and involve hydroxyl radicals and organic radicals.

Particulate pollutants have various natural and anthropogenic sources, and mainly originate from mineral dust and fuel combustion in motor vehicles. Since coal and petroleum often contain sulfur compounds, their combustion generates SO_2 , which is also a precursor to particulates in the atmosphere and acid rain. The many environmental and health effects of these air pollutants has been described by the World Health Organization in its air quality guideline [1].

It is known that atmospheric pollution is a very complex variable, which is affected by different chemical and physical individual pollutants. Researchers are making efforts to better analyze the pollutants and their relationships, to identify their sources, and to characterize their temporal and spatial patterns. The main objective of these works is to drag out useful results for the design of remediation strategies and reduction of emissions. Different multivariate approaches have been used to study the atmospheric pollution. One of the most popular methodology has been the principal component analysis (PCA), which has been utilized, for instance, to establish seasonal variations, dependence with meteorological conditions, or source apportionment studies [2–4]. However, an important drawback of PCA is the fact that factors do not often have straightforward interpretation [3]. Other alternative methods to PCA are the positive matrix factorization (PMF) [5], the multilinear engine (ME) [6], and multivariate curve resolution alternating least squares (MCR-ALS) [7], which have been utilized for the analysis of the factors contributing to atmospheric pollution, such as climatic variables (precipitation, air temperature, relative humidity, etc.) or different individual pollutants.

Recently, the Rasch model has been proposed as a technique to obtain a measure of atmospheric pollution considering several individual pollutants [8] and it has been also used to consolidate several measures of ozone levels into an overall variable to simplify the interpretation of ambient ozone in a medium-sized urban area [9]. With the formulation of the Rasch model as a measure technique [10], more information can be obtained than using the aforementioned methods. Thus, the influence of each individual pollutant on the environmental deterioration for a particular area, anomalies in any data at every sample location or relationships between different factors is analyzed using a probabilistic approach.

The Rasch model have been successfully used in many studies, proven their practical utility in different cases related to the environment [8,9,11–13]. As far as we know, there is no previous application of this useful methodology in this field. The present work is focused on extending and promoting the use of the Rasch model for the investigation and evaluation of atmospheric pollution in different locations, considering not only the main individual pollutants but also meteorological variables which exert an important influence on the contamination of the atmosphere. Some meteorological factors generate increased pollutants concentrations during episodes of high atmospheric stability and low wind speeds because of restricted dispersion [14]. It has been also reported that some pollutants exhibited a significative correlation with some climatic variables, usually positive with temperature [15] and inverse with precipitation [16]. However, other climate factors, as relative humidity and solar radiation, also affect some pollutant levels [2]. Moreover, stable atmospheric conditions increase the importance of urban-derived pollutants for urban air pollution levels, because of the low dispersion [17,18].

Although a Rasch analysis can also provide detailed evidence of anomalies (through the misfitting analysis) with respect to the operation of any particular variable which may over or under discriminate

relative to the summary discrimination of all variables, and anomalies with respect to the statistical independence of the variables, this analysis is not shown. Additional information about this issue can be obtained, for instance, in Moral et al. [9].

The objectives of this study are to: (1) Analyze the use of the Rasch model as a measurement tool to estimate with a rational basis a representative value of the atmospheric pollution in urban and rural sites; (2) utilize the Rasch methodology to find out how each individual pollutant and meteorological variable has an influence on the atmospheric pollution level; and (3) compare the pollution patterns in each location.

2. Materials and Methods

2.1. Study Sites

With the aim of analyzing different sites, the research was conducted in two urban and one rural locations in the Autonomous Community of Extremadura, in southwestern Spain (Figure 1). The urban sites are the two most populous cities of this region: Badajoz (around 150,000 inhabitants) and Cáceres (around 95,000 inhabitants). The rural site is the Monfragüe National Park. Extremadura (latitude between 37°57' and 40°29' N, longitude between 4°39' and 7°33' W) is one of the largest regions in Europe, with a surface area of approximately 41,600 km². Extremadura shows a great contrast, with wide agricultural and forest areas, and is considered to be one of the most important ecological enclaves in Europe. The main source of primary pollutants is the vehicle traffic because industry is not important in the region. Some small industrial activities can also contribute to air pollution.

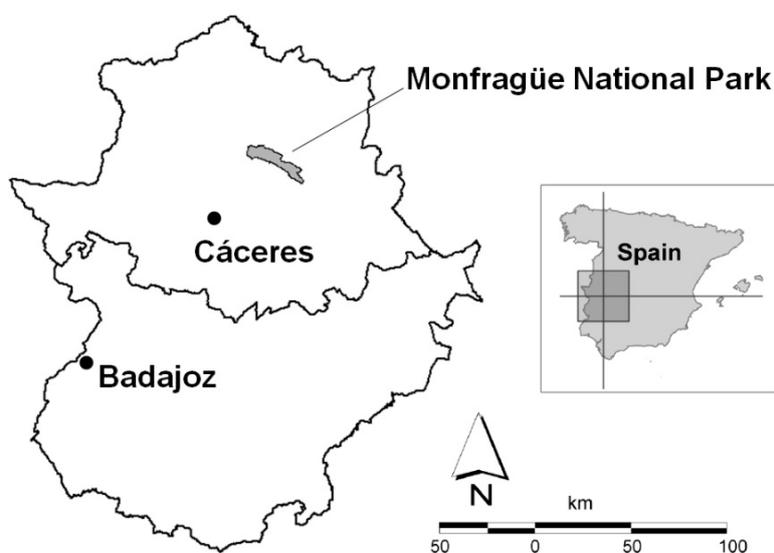


Figure 1. Location map of Badajoz, Cáceres and the Monfragüe National Park, with reference to the Autonomous Community of Extremadura and Spain.

The climate of Extremadura is characterized by a variation in both temperature and precipitation, typical of a Mediterranean climate. However, this feature is modified by the interior location of the region and by oceanic influences that penetrate the peninsula because of its proximity to the Atlantic. Mean annual precipitation reaches less than 600 mm in the majority of the areas of the region, even less than 400 mm in the center of the Guadiana valley, but it can reach much more than 1000 mm in the northern (Gredos) and eastern (Guadalupe) mountainous areas. One of the most important characteristics of the precipitation is its intra-annual variability. There is a dry season, from June to September, and a wet season, from October to May (80% of the precipitation falls between these months). Extremadura is a semiarid region, where the water balance is negative.

2.2. Data Collection and Treatment

Extremadura has six atmospheric pollution monitoring stations (five in medium and small-sized urban sites and one in a rural location), operated by the Department of Environment of the Extremadura Government, which are continuously measuring some air pollutants and meteorological variables (<http://xtr.gobex.es/repica/index.html>), providing measurements every minute. The procedures are established by Spanish legislation (Real Decreto 1073/2002). Pollutants measurements were saved in hourly period and data were collected from 1 January to 31 December 2016. Eight pollutants (NO_x , NO , NO_2 , O_3 , SO_2 , PM_{10} , CO , and benzene) and five climatic variables (precipitation, relative humidity, solar radiation, mean air temperature, and barometric pressure) were considered. Their mean daily value was computed and they were incorporated in a database. In consequence, the final data set is a matrix with 13 columns and 365 rows.

Next, it was necessary to obtain a measurement of the atmospheric pollution using different individual pollutants and climatic factors, with different units. This important problem may be solved with the application of a measured technique based on the Rasch model [19], using the WINSTEPS v. 3.69 computer program [20]. To do that, first, a transformation of the pollutants and the climatic factor measures to common categories was performed and data were incorporated in the cells of the matrix. Five categories or levels were established; in consequence, measures were categorically coded according to a plan where each pollutant and climatic variable was rated on a scale (1–5) for each day. The minimum and maximum values of the scale were assigned to the maximum and minimum values of each pollutant and climatic factor and other intermediate values for all different days were obtained through interpolation. Thus, a measure assigned to level 1 indicates the lowest contribution to atmospheric pollution and, on the contrary, a measure assigned to level 5 indicates the highest contribution to atmospheric pollution. With 13 pollutants and climatic factors taken into account, the highest possible raw score for the days in 2016 is 65 (the most potentially polluted) and the lowest possible score is 13 (the least potentially polluted). Considering all days, the highest possible raw score for pollutants and climatic factors is 1825 (the most influential on atmospheric pollution) and the lowest possible score is 365 (the least influential on atmospheric pollution).

With respect to the data distribution, the Rasch model has certain distinctive and interrelated features [10]: it is concerned principally with the measurement of individuals (considered days in this study), rather than with distributions among populations, and it is concerned with establishing a basis for meeting a priori requirements for measurement and, consequently, do not invoke any assumptions about the distribution of levels of a trait in a population.

The main outputs of the program are explained in this work; the empirical hierarchy of the pollutants and climatic factors, which is illustrated using variable maps and related to all days, with each reported in logits, the statistics to show how well the data fit the model, and, moreover, the way to study the unexpected values.

2.3. The Rasch Model

The Rasch probabilistic model is well known for its efficiency and precision of transforming categorical item responses to objective scale measures. Moreover, it also has an interesting capacity to consolidate data that is already reported sometimes in several scale metrics. This model can synthesize and consolidate seemingly disparate data into a uniform analytical framework. In this case study, the purpose of this procedure is to transcend several heterogeneous measures and consolidate them into an overall variable that simplifies interpretation of air pollution exposure.

The data are arranged in matrix form, where the rows are the days, in 2016, and the columns the pollutants and climatic factors, and each cell reflects the category. In consequence, one possible way of obtaining a ranking is summing the categories by rows or by columns, i.e., to sum the categories of all pollutants and climatic factors for each day, and of all the days for each pollutant and climatic factor. However, these sums establish separate rankings for the days and the pollutants and climatic factors,

and the procedure does not discriminate between ranking days in terms of pollutants and climatic factors and, conversely, these in terms of days.

The Rasch model uses the traditional sum of the item ratings as a starting point for estimating response probabilities. This model is based on the simple idea that some items, in this case study pollutants and climatic factors, are more important to subjects, in this case days, than other items. The Rasch model constructs a line of measurement with the items placed hierarchically on this line according to their importance to subjects and the validity of a given test is assessed through examination of this item ordering, evaluating whether all items work together to measure a single overall variable.

In order to determine how well each item contributes to the atmospheric pollution measurement, chi-square fit statistics, known as infit and outfit mean-square (Infit and Outfit MNSQ), ratios of observed residual variance to expected residual variance, where expectations are 1, should be computed. Infit is an information-weighted or inlier-sensitive fit statistic that focuses on the overall performance of an item or subject, and outfit is an outlier-sensitive fit statistic that picks up rare events that have occurred in an unexpected way. Usually, items that fall between the infit and outfit limits of 0.6 and 1.5 are accepted and those with values beyond these thresholds are removed [19,20].

Some previous works [8,11] can be revised to obtain more information about the mathematical formulation of the Rasch model. The different contribution of the eight pollutants and five climatic variables, previously indicated, to determine a measure of atmospheric pollution at each location was achieved through the stages shown in Figure 2.

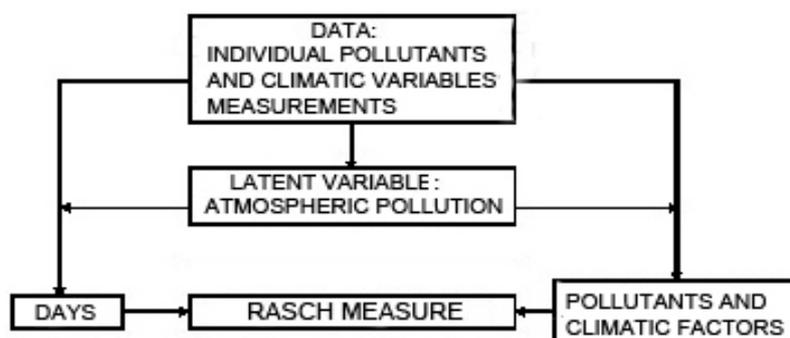


Figure 2. Schematic diagram of the stages involved in the formulation of the Rasch model.

3. Results and Discussions

3.1. Data Response to the Model at Each Site

After processing the matrix of categorical values by the WINSTEPS program, the first information to be considered is if the data fit the model reasonably by analyzing the infit and outfit statistics. Consequently, according to the infit and outfit MNSQ values contained in Table 1 (between 1.09 and 1.23), there is a clear evidence about the agreement between the data and the model. Moreover, the mean standardized (ZSTD) infit and outfit, which are the sum of squares standardized residuals given as a Z-statistics [21], are expected to be 0; in this case study, they both are near zero for samples (days) at the three locations (Table 2). However, when items are considered, better results are obtained for Monfragüe, with values near zero, than for Badajoz or Cáceres (Table 1), denoting that the data fit the model better in the rural site than in the urban locations, despite in all cases the overall fit is acceptable. The same is apparent when the standard deviation of the infit MNSQ is analyzed for days. This is also an index of the overall misfit and a value below 2 is considered acceptable [22]. There are no important misfits in this case study because their values are 0.63, 0.74, and 0.81 (Table 2), also indicating an acceptable overall fit of the data.

Table 1. Overall model fit information; summary of all 13 individual pollutants and climatic variables (items) for the three locations: sum of points of the common scale (total score), days or items taken into account (count), logit position of the days and items along the straight line that represents the latent variable, atmospheric pollution (measure), and standard error of measurement (model error). Infit and outfit MNSQ are mean-square fit statistics to verify if items fit the model; infit and outfit ZSTD are standardized fit statistics to verify if items fit the model.

	Total Score	Count	Measure	Model Error	Infit		Outfit		
					MNSQ	ZSTD	MNSQ	ZSTD	
Badajoz									
Mean	788.4	332	0.00	0.10	1.23	−1.4	1.09	−1.4	
Standard Deviation	336.1	0	1.70	0.10	0.78	5.6	0.62	5.1	
Maximum	1633.0	332	3.88	0.41	2.93	9.9	2.84	9.9	
Minimum	338.0	332	−3.88	0.05	0.47	−9.9	0.49	−9.3	
Cáceres									
Mean	699.1	304	0.00	0.08	1.23	−1.1	1.20	−0.9	
Standard Deviation	305.3	0	1.36	0.05	0.79	5.3	0.87	4.7	
Maximum	1488.0	304	2.57	0.22	3.15	9.9	3.79	9.9	
Minimum	325.0	304	−3.27	0.05	0.46	−8.7	0.50	−7.9	
Monfragüe									
Mean	772.8	333	0.00	0.10	1.16	−0.2	1.09	−0.3	
Standard Deviation	380.8	0	1.81	0.06	0.48	5.4	0.48	5.2	
Maximum	1619.0	333	2.65	0.22	1.91	9.9	1.98	9.9	
Minimum	355.0	333	−3.72	0.05	0.53	−8.6	0.54	−8.0	

Table 2. Overall model fit information; summary of all 365 days in 2016 for the three locations: sum of points of the common scale (total score), days or items taken into account (count), logit position of the days, and items along the straight line that represents the latent variable, atmospheric pollution (measure), and standard error of measurement (model error). Infit and outfit MNSQ are mean-square fit statistics to verify if items fit the model; infit and outfit ZSTD are standardized fit statistics to verify if items fit the model.

	Total Score	Count	Measure	Model Error	Infit		Outfit		
					MNSQ	ZSTD	MNSQ	ZSTD	
Badajoz									
Mean	30.9	13	−0.75	0.31	1.01	−0.1	1.03	0.1	
Standard Deviation	5.7	0	0.53	0.05	0.63	1.1	1.29	0.9	
Maximum	44.0	13	0.32	0.68	6.40	4.7	9.90	7.8	
Minimum	17.0	13	−3.18	0.27	0.26	−2.4	0.23	−1.1	
Cáceres									
Mean	29.9	13	−0.68	0.30	1.01	−0.2	1.16	0.1	
Standard Deviation	5.1	0	0.50	0.06	0.74	1.3	1.40	1.3	
Maximum	42.0	13	0.24	0.59	5.51	5.0	9.90	7.5	
Minimum	17.0	13	−2.81	0.26	0.15	−2.8	0.18	−1.6	
Monfragüe									
Mean	30.2	13	−1.04	0.34	1.01	−0.2	1.07	0.0	
Standard Deviation	5.1	0	0.59	0.04	0.81	1.4	1.33	1.2	
Maximum	41.0	13	0.14	0.56	7.04	5.6	9.90	6.7	
Minimum	18.0	13	−3.02	0.31	0.12	−3.2	0.13	−2.0	

The next step is to analyze the internal consistency of data, in the sense that expected results with the Rasch model will be accurate if measures of the items are properly performed, without significant errors. Thus, there is a reliability statistics, which is expected to be close to 1. Acceptable values would be over 0.7 [23]. In this study, reliability of days was 0.66 for Badajoz and Monfragüe, and 0.61 for Cáceres; for pollutants and climatic factors, they were 0.99 for all locations. In spite of computing

statistics lower than the suggested limit for days, they are very close to the recommended values; moreover, the fact that item reliabilities are very high is indicative of the adequate consistency of data, and probably measures do not have significant errors.

It is also necessary to verify how the assignment scale has been utilized [20]. There is no general rule to initially define the correct number of categories. Following the experience in previous works [12,13], five categories were chosen. After analyzing some parameters, it was clear that this number was appropriate for the three locations. Thus, according to Table 3, the “observed average” and the “structure calibration” increase by category value, all infit and outfit MNSQ values are between 0.6 and 1.5, and the “observed average” values are very close to the “sample expected” ones, as it is recommended [22].

Table 3. Response scale use: number of times the category was selected considering all days and items, pollutants and climatic factors (observed count), mean value of logit positions modelled in the category (observed average), optimum values of the average logit positions for the data (sample expected), and logit calibrated difficulty of the step representing the transition points between one category and the next (structure calibration). Infit and outfit MNSQ are mean-square fit statistics to verify if items fit the model.

Category	Observed Count (%)	Observed Average	Sample Expected	Infit MNSQ	Outfit MNSQ	Structure Calibration
Badajoz						
1	41	-1.97	-2.00	1.11	1.06	None
2	20	-0.82	-0.73	0.74	0.93	-0.47
3	14	-0.30	0.28	1.04	1.14	-0.16
4	10	0.20	0.16	0.71	0.63	0.25
5	15	1.68	1.66	1.08	1.76	0.39
Cáceres						
1	43	-1.66	-1.64	0.96	0.98	None
2	22	-0.66	-0.71	0.80	0.65	-0.47
3	12	-0.25	-0.26	0.95	1.29	0.16
4	8	0.09	0.15	0.86	0.88	0.39
5	15	1.30	1.30	1.11	2.19	0.48
Monfragüe						
1	45	-2.52	-2.53	1.01	1.03	None
2	19	-0.88	-0.89	1.00	0.97	-0.83
3	11	-0.23	-0.09	0.99	1.14	0.10
4	9	0.45	0.45	0.73	0.94	0.31
5	16	1.52	1.48	1.01	1.46	0.41

3.2. Analysis of the Rasch Measure: Atmospheric Pollution

One of the main outputs of the Rasch model is the denominated variable map. In this particular case, all pollutants and climatic variables and days are displayed in the same scale (Figure 3). In consequence, the relative distribution of the days is provided in the upper half of the continuum, according to the associated atmospheric pollution, which has been achieved by means of the 13 pollutants and climatic variables taken into account, and, similarly, these items are provided in the lower half of the diagram, classified according to the influence on atmospheric pollution.

The item that obtained the highest measure in all three studied locations, and is to the right in the continuum (Figure 3), is the CO (their measures were 3.88, 2.57, and 2.65, in Badajoz, Cáceres and Monfragüe, respectively; see Table 4). This means it is the less common item, i.e., its influence on atmospheric pollution is the lowest. The same is apparent in Table 4 because its raw score is consequently the lowest. At the other extreme, the lack of precipitation (it is represented as rainfall in Figure 3, Tables 4 and 5) is situated (their measures were -3.88, -3.27, and -3.72, in Badajoz, Cáceres and Monfragüe, respectively; see Table 4), with the highest score, being the more common item because

this climatic variable affects all days, i.e., it is the most influential item on the atmospheric pollution in all locations.

After analyzing the variable maps (Figure 3) and the distribution of all items, important differences are apparent between the urban sites and the rural one. As it could be expected, because it is a National Park, Monfragüe is not affected by many individual pollutants. Only ozone has an important impact at this location, and SO₂ has a minor influence (probably generated from processes of organic decomposition, although the concentrations were always very low). With respect to the climatic variables, besides the lack of precipitation (which is the most influential factor on atmospheric pollution as it was previously indicated), the other four factors are important during many days, and they are close to tropospheric ozone, indicating their influence on this pollutant.

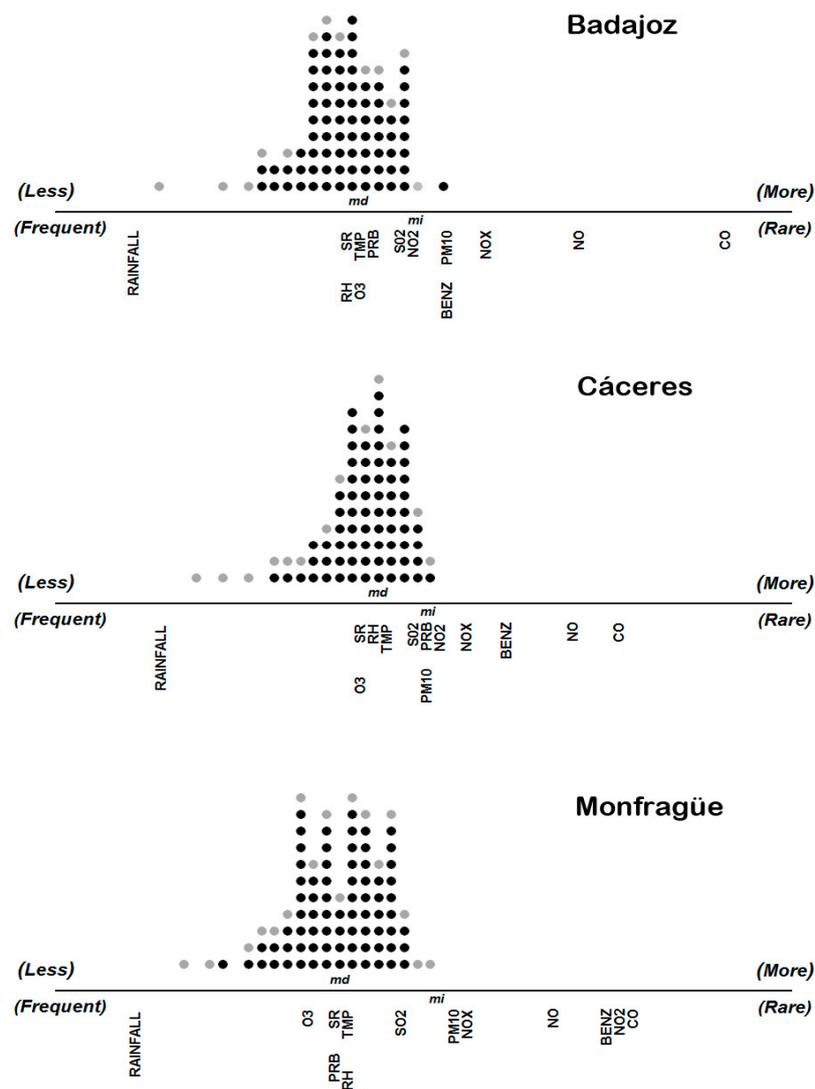


Figure 3. Variable maps for the three locations. The straight line represents the latent variable: atmospheric pollution. Days in 2016 (each black point is 4 days and each grey point is 1 to 3 days) are represented above the line according to their atmospheric pollution measure: to the right those when higher atmospheric pollution level occurred; to the left those when lower atmospheric pollution level occurred. Individual pollutants and climatic factors are below the line: to the right those with lower influence on atmospheric pollution (rare); to the left those with higher influence on atmospheric pollution (frequent). Rainfall represents the lack of precipitation, RH is the relative humidity, SR is the solar radiation, TMP is the mean air temperature, PRB is the barometric pressure, and BENZ is benzene.

Table 4. Influence of each pollutant and climatic factor on the atmospheric pollution. Total score is the sum of the points of the common scale for each item considering all days in 2016; measure is the atmospheric pollution level.

Item	Badajoz		Cáceres		Monfragüe	
	Total Score	Measure	Total Score	Measure	Total Score	Measure
CO	338	3.88	325	2.57	355	2.65
NO	372	2.03	345	1.91	390	1.68
NO _x	476	0.85	479	0.53	520	0.43
Benzene	577	0.36	418	0.92	363	2.34
PM ₁₀	585	0.33	623	-0.02	534	0.36
NO ₂	674	0.04	568	0.15	359	2.48
SO ₂	745	-0.16	654	-0.11	697	-0.29
Barometric pressure	863	-0.45	604	0.03	1042	-1.18
Temperature	964	-0.68	839	-0.56	979	-1.03
O ₃	991	-0.74	941	-0.78	1154	-1.45
Relative humidity	1010	-0.78	867	-0.62	989	-1.06
Solar radiation	1021	-0.81	937	-0.77	1046	-1.19
Rainfall	1633	-3.88	1488	-3.27	1619	-3.72

Table 5. Guttman scalograms for all individual pollutants and climatic factors (13) and days in 2016 considered. Only some days, those with higher and lower atmospheric pollution level, are shown for each location.

Day	Rainfall	Solar Radiation	Relative Humidity	O ₃	Temperature	Barometric Pressure	SO ₂	NO ₂	PM ₁₀	Benzene	NO _x	NO	CO
Badajoz													
159	5	5	5	5	5	1	3	3	4	4	2	1	1
206	5	5	5	5	5	5	2	4	4	1	1	1	1
237	5	5	4	5	5	5	2	5	3	1	2	1	1
...
18	5	1	1	1	1	1	2	1	1	2	1	1	1
132	1	2	1	2	3	1	1	1	1	1	1	1	1
Cáceres													
25	5	3	2	3	1	5	3	1	5	4	3	3	4
200	5	5	5	5	5	3	4	1	3	2	1	1	1
231	5	3	3	4	3	2	5	5	4	3	2	1	1
...
68	3	2	1	1	2	2	1	1	1	1	1	1	1
73	3	2	1	1	2	1	1	1	1	1	1	1	1
Monfragüe													
206	5	5	4	5	5	5	3	2	2	1	1	2	1
207	5	4	4	5	5	5	3	3	2	1	1	2	1
261	5	4	3	5	4	3	3	3	4	4	1	1	1
...
38	3	3	1	1	1	1	1	1	2	1	1	1	1
42	1	3	1	2	1	2	2	1	1	1	1	1	1

In both urban locations, ozone is the most influential pollutant on the environmental deterioration. Similarly to Monfragüe, the climatic variables which influence ozone generation are around it, denoting their relationship. However, some others individual pollutants exert a limited influence on urban atmospheric pollution, as SO₂, PM₁₀, and NO₂, mainly produced from motor vehicles and other fuel combustion processes.

In general, it was apparent that the considered climatic factors (except precipitation) are aggregated, close to tropospheric ozone, and the individual pollutants are distributed along the line. However, some of them, located to the right in the continuum, have very low score, denoting their low influence on the atmospheric pollution. In consequence, a ranking of all the considered items have been obtained as an output of the Rasch model. According to the order established after processing all data, CO is the pollutant with higher measure, followed by NO, NO_x, and benzene in the urban sites, and NO₂,

benzene, NO, NO_x, and PM₁₀ in the rural location, but their influence on atmospheric pollution is lower than the other pollutants and climatic factors.

Another ranking of all days according to their Rasch measure can be obtained. Figure 3 displays a continuous distribution of days, with most of them aggregated. However, some days, located to the left in the continuum, have very low score, suggesting their low atmospheric pollution. Conversely, the days located to the right in the continuum are those in which high atmospheric pollution occurred.

It is important to denote that the pollutants levels at the three locations were usually very low and consequently, the atmospheric pollution level was also very low. This fact is reflected in the variable maps: the mean Rasch measure for days is lower than the mean Rasch measure for pollutants and climatic factors (Figure 3). The separation between both Rasch measures is greater at the rural site, Monfragüe, than at the urban locations, and, in turn, the separation was greater at Badajoz than at Cáceres (see Tables 1 and 2). That means Cáceres is the location where there were more days in which the probability to find high atmospheric pollution was more important, followed by Badajoz (despite this city has more population and, consequently, the number of motor vehicles is considerably higher), and as it could be expected, Monfragüe has very low probability to find high atmospheric pollution.

Figure 4 shows the mean atmospheric pollution level, the Rasch measure, for each month during 2016. Higher values are evident for both urban locations and, in turn, the Rasch measure was considerably higher in Cáceres during many months, basically in spring and summer, when atmospheric pollution was more important because of the increasing contribution of the ground-level ozone. During summer, atmospheric pollution levels in the three sites were more similar, coinciding with maximum solar radiation, mean air temperature, and, in consequence, higher contribution of the tropospheric ozone. In contrast, in winter, when the ozone levels are lower, other individual pollutants generated in urban areas, as NO₂, PM₁₀, or benzene, acquire a highest relative level, because of the increased emissions from heatings or heavier traffic in cities compared to the vacation period, as it has been also observed in other locations [24]; in consequence, atmospheric pollution levels are considerably higher in the urban sites than in the rural one.

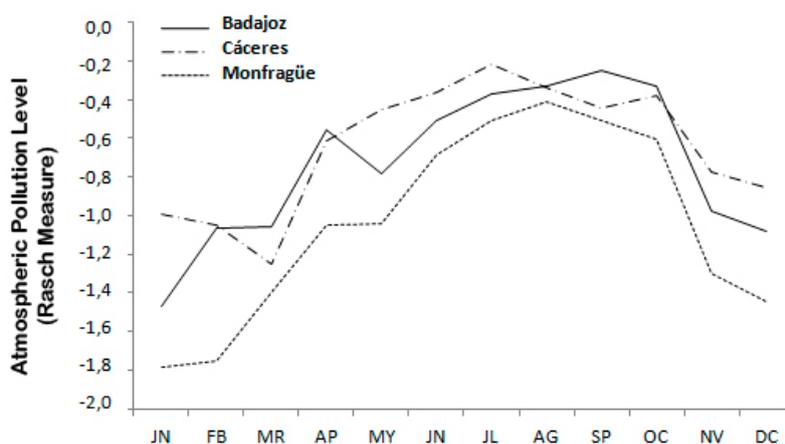


Figure 4. Temporal evolution of atmospheric pollution levels at the two urban (Badajoz and Cáceres) and the one rural (Monfragüe) locations, in 2016.

As it was previously mentioned, ozone is the only individual pollutant which exerts an important influence during a large number of days in Monfragüe, and, additionally, this number of days is greater than in Badajoz or Cáceres (Figure 3). The existence of higher ozone concentrations in rural areas, with low anthropogenic activities, than in urban areas has been also reported in other regions [25]. One of the main reasons of this fact is the impossibility to generate the titration effect of NO emissions because this ozone precursor is not as important in rural areas as it is in urban ones [24]. Moreover, although most emissions of ambient air pollutants are from local or regional sources, under certain atmospheric conditions ozone precursors can travel long distances. Thus, in this case, Madrid, the most important

city in Spain, is located around 250 km to the east of Monfragüe, and, in its surroundings there is one of the most important industrial areas of Spain. Emissions of NO_x into the atmosphere from a variety of anthropogenic sources, mainly traffic and industrial activities, from the area around Madrid can reach the Monfragüe Park in some days. In consequence, many of the ozone episodes in this area occur with a wind flow from the southwesterly direction coming from a region with strong emissions of NO_x as Madrid.

Another useful output from the Rasch model is the Guttman scalogram [26]. It is a complementary tool to display all days according to their level of atmospheric pollution. As it can be seen in Table 5, days are sorted in a descending order, by their level of atmospheric pollution. Simultaneously, the Guttman scalogram shows all individual pollutants and climatic factors arranged in the order indicated in the first row (Table 5), enabling to show their influence on the latent variable during each day.

The Guttman scalogram has the advantage that a single variable is considered to analyze the individual behavior of each day. In the same way, the individual pattern of each individual pollutant or climatic factor can be studied. Pollutants and climatic factors located to the left in Table 5 had higher influence on atmospheric pollution than those located to the right. Thus, a hierarchical order is obtained and the unexpected scores for some items can be easily visualized. For instance, CO, NO_x, benzene, and NO scores are usually low for the days with higher atmospheric pollution in Monfragüe (Table 5). However, an unexpected high score for benzene on day 261 was detected. Inversely, day 38, located in the bottom of the ranking, where scores of pollutants are very low, had an unexpected high value for SO₂.

Any other days can be analyzed in a similar way, detecting the particular item in which a low or high value exists with respect to the pattern of the surrounding days. For example, in Monfragüe, day 206 (19 August), which is placed at the top of the scalogram (Table 5), had the highest scores for some items (precipitation, ozone, relative humidity, solar radiation, mean air temperature, and barometric pressure); it was the day when the highest atmospheric pollution occurred. If it is compared with days 207 (20 August) and 261 (19 October), which also had the maximum total score but are the second and third in the ranking, it is observed that they all had the maximum score for two items, precipitation and barometric pressure, but for the other items the scores were different. Days 207 and 261 had a lower score in ozone and solar radiation, two of the main items which affect atmospheric pollution in Monfragüe; consequently, they are after day 206. Those days at the top in Table 4 had high atmospheric pollution (higher Rasch measure). In contrast, those days at the bottom had low scores for all items; they were the days with the lowest atmospheric pollution (lower Rasch measure). The last in the Guttman scalogram is day 42 (19 February); in consequence, it was the day when the lowest atmospheric pollution took place.

In Badajoz, according to the Guttman scalogram, the days when the highest atmospheric pollution was generated had higher scores for the same items (precipitation, ozone, relative humidity, solar radiation, mean air temperature, and barometric pressure) than the rural site (Monfragüe), but the participation of SO₂ and NO₂ in the environmental deterioration was also important during some days, usually located at the top of the scalogram, with high scores.

In the other studied urban location, Cáceres, the days with highest atmospheric pollution correspond to the highest scores for some of the same relevant items in Badajoz (precipitation, ozone, relative humidity, solar radiation, and mean air temperature), under the influence of barometric pressure, SO₂ and PM₁₀ are important factors for the atmospheric pollution level during some of the days situated at the top of the scalogram. Curiously, day 25 (25 January) was when the highest atmospheric pollution occurred, despite the relatively low scores in some of the main items, particularly the mean temperature and solar radiation (and, in consequence, ozone), but high scores in other individual pollutants as NO, NO_x, and CO which had usually very low scores during the year.

4. Conclusions

The successful formulation of the Rasch model to define a measure of atmospheric pollution, integrating different measurements of individual pollutants (CO, SO₂, NO_x, NO₂, NO, O₃, PM₁₀,

and benzene) and climatic factors (precipitation, relative humidity, solar radiation, mean air temperature, and barometric pressure), is the novel aspect of this work. Moreover, this probabilistic and objective model, based on the data, can detect the influence of each pollutant or climatic variable on the environmental deterioration and the pollution measurement for each day.

After applying the Rasch method, it was highlighted how climatic variables are the main items affecting the atmospheric pollution in the three study locations, with the lack of precipitation being the most important factor. Most of the considered individual pollutants do not influence significantly the atmospheric pollution, which is probably due to the low levels measured in the locations. Only ozone is important in all sites and, particularly, in the rural one, despite it is a National Park. In consequence, measures to reduce air pollution in both cities, Badajoz and Cáceres, should be addressed to reduce the emissions of ozone precursors from traffic, restricting the use of vehicles with internal combustion engines and promoting the use of less polluting vehicles such as those with electric motors. In the Monfragüe Park, the transport of ozone precursors from other areas made it difficult to propose local measures to reduce the episodes of air pollution.

An additional output of the Rasch model is the Guttman scalogram, which constitutes a useful tool to detect unexpected values of any pollutant or climatic factor during each day, or, inversely, the individual pattern of each pollutant or climatic factor during all considered days. Usually these unexpected values are due to local transitory circumstances, such as episodes of intense rainfall or, in contrast, the occurrence of days with Saharan dust intrusion that occurs in summer, increasing the level of particulate matter.

The use of the Rasch model as a measurement tool constitutes a logical and objective way to define a coherent variable (atmospheric pollution level as the Rasch measure) and it is a powerful tool to better know the interrelationships between all different variables or factors which can affect complex processes as the dynamics of environmental contamination. Since georeferenced information was used, it can be implemented in a geographical information system (GIS) to generate different types of maps. In the future, the combination of the Rasch model and GIS capabilities can be a powerful tool to develop an appropriate environmental and managing policy.

Author Contributions: F.J.M. conceptualized, collected, and analyzed the data, performed the analyses, and wrote the manuscript. F.J.R. also performed the analyses and validation. P.V. and F.L. collected the data and optimized the manuscript.

Funding: This research was funded by the Junta de Extremadura and the European Regional Development Fund (ERDF) through the projects GR18086 and GR18081 linked to the VI Regional Plan for Research, Technological Development and Innovation from the General Government of Extremadura 2017–2020. We also acknowledge the support received from the Air Quality Surveillance Network of Extremadura (REPICA), project 1855999FD022, partially financed by the European Regional Development Fund (ERDF).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. Air Quality Guidelines—Global Update 2005. In *Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide*; World Health Organization, Regional Office for Europe: Copenhagen, Denmark, 2006.
2. González Gallero, F.J.; Galán Vallejo, M.; Umbría, A.; Baena Gervilla, J. Multivariate statistical analysis of meteorological and air pollution data in the Campo de Gibraltar region, Spain. *Environ. Monit. Assess.* **2006**, *119*, 405–423. [[CrossRef](#)] [[PubMed](#)]
3. Shah, M.H.; Shaheen, N. Statistical analysis of atmospheric trace metals and particulate fractions in Islamabad, Pakistan. *J. Hazard. Mater.* **2007**, *147*, 759–767. [[CrossRef](#)] [[PubMed](#)]
4. Chang, S.C.; Lee, C.T. Evaluation of the temporal variations of air quality in Taipei City, Taiwan, from 1994 to 2003. *J. Environ. Manag.* **2008**, *86*, 627–635. [[CrossRef](#)] [[PubMed](#)]
5. Paatero, P. A weighted non-negative least squares algorithm for three-way ‘PARAFAC’ factor analysis. *Intell. Lab. Syst.* **1997**, *38*, 223–242. [[CrossRef](#)]
6. Paatero, P. The multilinear engine: A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model. *J. Comput. Graph. Stat.* **1999**, *8*, 854–888.

7. Felipe-Sotelo, M.; Gustems, L.; Hernández, I.; Tauler, R. Investigation of geographical and temporal distribution of tropospheric ozone in Catalonia (North-East Spain) during the period 2000–2004 using multivariate data analysis methods. *Atmos. Environ.* **2006**, *40*, 7421–7436. [[CrossRef](#)]
8. Moral, F.J.; Álvarez, P.; Canito, J.L. Mapping and hazard assessment of atmospheric pollution in a medium sized urban area using the Rasch model and geostatistics techniques. *Atmos. Environ.* **2006**, *40*, 1408–1418. [[CrossRef](#)]
9. Moral, F.J.; Rebollo, F.J.; Valiente, P.; López, F.; Muñoz de la Peña, A. Modelling ambient ozone in an urban area using an objective model and geostatistical algorithms. *Atmos. Environ.* **2012**, *63*, 86–93. [[CrossRef](#)]
10. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Revised and Expanded ed.; University of Chicago: Chicago, IL, USA, 1980.
11. Moral, F.J.; Terrón, J.M.; Rebollo, F.J. Site-specific management zones based on the Rasch model and geostatistical techniques. *Comp. Electron. Agric.* **2011**, *75*, 223–230. [[CrossRef](#)]
12. Moral, F.J.; Rebollo, F.J.; Terrón, J.M. Analysis of soil fertility and its anomalies using an objective model. *J. Plant Nutr. Soil Sci.* **2012**, *175*, 912–919. [[CrossRef](#)]
13. Moral, F.J.; Rebollo, F.J.; Méndez, F. Using an objective model to estimate overall ozone levels at different urban locations. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 455–465. [[CrossRef](#)]
14. He, J.; Gong, S.; Yu, Y.; Yu, L.; Wu, L.; Mao, H.; Song, C.; Zhao, S.; Liu, H.; Li, X.; et al. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environ. Pollut.* **2017**, *223*, 484–496. [[CrossRef](#)] [[PubMed](#)]
15. Kalisa, E.; Fadlallah, S.; Amani, M.; Nahayo, L.; Habiyaemye, G. Temperature and air pollution relationship during heatwaves in Birmingham, UK. *Sustain. Cities Soc.* **2018**, *43*, 111–120. [[CrossRef](#)]
16. Feng, H.; Zou, B.; Wang, J.; Gu, X. Dominant variables of global air pollution-climate interaction: Geographic insight. *Ecol. Indic.* **2019**, *99*, 251–260. [[CrossRef](#)]
17. Arnfield, A.J. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *Int. J. Climatol.* **2003**, *23*, 1–26. [[CrossRef](#)]
18. Papanastasiou, D.K.; Melas, D. Climatology and impact of air quality of sea breeze in an urban coastal environment. *Int. J. Climatol.* **2009**, *29*, 305–315. [[CrossRef](#)]
19. Álvarez, P. *Several Noncategorical Measures Define Air Pollution Construct*. *Rasch Measurement in Health Science*; Jam press: Maple Grove, MN, USA, 2005.
20. Linacre, J.M. *WINSTEPS (Version 3.69) [Computer Program]*; Winsteps.com: Chicago, IL, USA, 2009.
21. Edwards, A.; Alcock, L. Using Rasch analysis to identify uncharacteristic responses to undergraduate assessments. *Teach. Math. Its Appl.* **2010**, *29*, 165–175. [[CrossRef](#)]
22. Bode, R.K.; Wright, B.D. Rasch measurement in higher education. In *Higher Education: Handbook of Theory and Research*, vol. XIV; Smart, J.C., Tierney, W.G., Eds.; Agathon Press: New York, NY, USA, 1999.
23. Sekaran, U. *Research Methods for Business: A Skill Building Approach*; John Wiley & Sons Inc.: Singapore, 2000.
24. Alier, M.; Felipe, M.; Hernández, I.; Tauler, R. Trilinearity and component interaction constraints in the multivariate curve resolution investigation of NO and O₃ pollution in Barcelona. *Anal. Bioanal. Chem.* **2011**, *399*, 2015–2029. [[CrossRef](#)] [[PubMed](#)]
25. Li, P.; De Marco, A.; Feng, Z.; Anav, A.; Zhou, D.; Paoletti, E. Nationwide ground-level ozone measurements in China suggest serious risks to forests. *Environ. Pollut.* **2018**, *237*, 803–813. [[CrossRef](#)] [[PubMed](#)]
26. Tristán, A. *Análisis de Rasch Para Todos*; Ceneval: México city, Mexico, 2002.

